Metadata Extraction for EARS

Greg Sanders
for the NIST Gang

May 9, 2002
EARS Workshop

Summary of MDE Experiment

- Intent: Get proposed metadata types from research groups
  - Definition of each metadata type
  - Examples of use in the data provided
- We provided 3 excerpts of each type of data
  - Switchboard
  - Broadcast news
  - Meeting room
- Result: Participants proposed a wide variety of annotations, defined each, and used them to mark up the example data

Four classes of the metadata proposed in the experiment

- Identifying speakers
- Making transcripts more easily understandable
- Disambiguating transcriptions
- Marking acoustic phenomena

NIST National Institute of Standards and Technology

Identifying Speakers

- Speaker change detection
- Gender of speaker
  - We see four reasonable values:
    - Male / Female / Child / Unknown
- Speaker ID
  - Could be an index into a table of speaker info

NIST National Institute of Standards and Technology

Understanding Transcriptions More Easily

- Sentence boundaries
  - Difficult research problem for spontaneous speech
- Verbal edit (also called an edit-interval)
- Acronyms (with expansion)
- URL reference
- Named entities/proper nouns, permits capitalization
- Numeric expressions, format as numeric
- Temporal expressions, format as date, time, etc.

NIST National Institute of Standards and Technology

Disambiguating Transcriptions

- Comma

  Information probably from prosody or pauses

  He said<comma /> they are fools.

  In the above example, we disambiguate a direct quote from an indirect quote.

  The police found lots of stuff in the thief's garage<comma /> plastic deer<comma /> horns<comma /> and other things stolen from the hunting club.

  In the above example, we disambiguate "plastic deer" from "plastic deer horns"

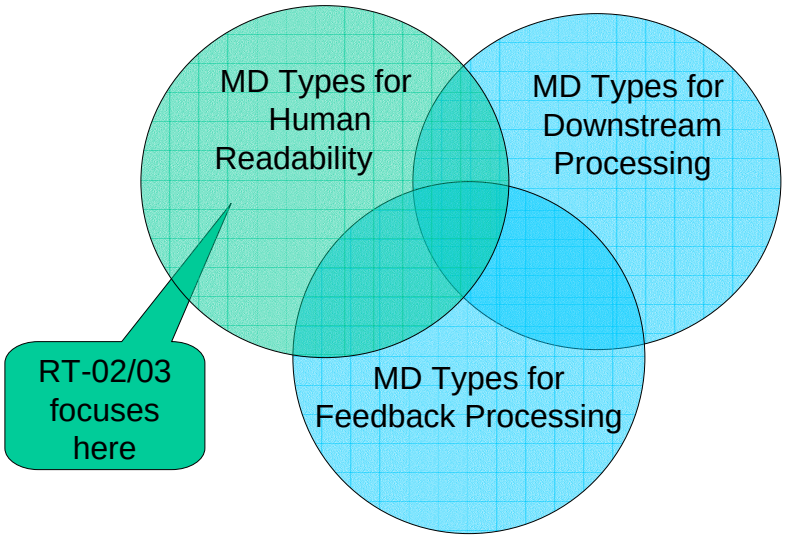
NIST National Institute of Standards and Technology

Marking Acoustic Phenomena

- Momentary phenomena
  - Noises
  - Non-speech vocal sounds by the speaker(s)
- Sustained phenomena
  - Background speech ("accompanying" at least one transcribed word)
  - Background non-speech ("accompanying" at least one transcribed word)

NIST National Institute of Standards and Technology

Goals for EARS MDE



NIST National Institute of Standards and Technology

MDE Types for Human Readability

- Obvious types of interest for RT-02 / RT-03:
  - Speaker change detection/identification (RT-02)
    - permits association of speakers with orthography
  - Sentence boundary detection/classification
    - permits natural orthographic segmentation, capitalization, and punctuation
    - VERY difficult for spontaneous speech, will require research
  - Acronym detection and expansion
    - permits capitalization of acronyms and pointer to expansion
  - Verbal edit detection
    - permits removal of verbal edits for cleaned up transcript
  - Named entity/proper noun detection/classification
    - permits capitalization of proper nouns
  - Numeric expression detection/classification
    - permits numeric representation of numbers
  - Temporal expression detection/classification
    - permits natural representation of time/date expressions

NIST National Institute of Standards and Technology

Example of a Dramatic Script

```
EURIPIDES Why, at first starting here's a fault skyhigh.
AESCHYLUS (to DIONYSUS) You see your folly?
DIONYSUS Have your way; I care not.
AESCHYLUS (to EURIPIDES) What is my fault?
EURIPIDES Begin the lines again.
AESCHYLUS "Grave Hermes, witnessing a father's power-“
EURIPIDES And this beside his murdered father's grave Orestes speaks?
AESCHYLUS I say not otherwise.
EURIPIDES Then does he mean that when his father fell By craft and
   violence at a woman's hand, The god of craft was witnessing the
   deed?
AESCHYLUS It was not he: it was the Helper Hermes He called the grave:
   and this he showed by adding It was his sire's prerogative he held.
EURIPIDES Why this is worse than all. If from his father He held this
   office grave, why then-
DIONYSUS He was a graveyard rifler on his father's side.
AESCHYLUS Bacchus, the wine you drink is stale and fusty.
DIONYSUS Give him another: (to EURIPIDES) you, look out for faults.
```

NIST National Institute of Standards and Technology

Later in EARS and Other Programs

- Metadata supporting downstream applications
  - Information analysis, information extraction, etc.
  - Topic/concept tracking
  - Translation
  - Summarization
- Video (outside scope of EARS)
  - Gestures
  - Facial displays and gaze direction that implement turn-taking behaviors
  - Shared visual context -- whiteboard, handouts, . . .
- Metadata intended to feed back into ASR
  - Background speech and noise/music

NIST National Institute of Standards and Technology

Discussion

- Are dramatic scripts a good prototype model for **formatting** the transcriptions?
- **What** are the right metadata **types** for near term?
  - Comments on MDE types for human readability?
- **What** should be **evaluated** in the future and **when**?
  - RT-03: Types for human readability
  - RT-04: ?
  - RT-05: ?
  - RT-06: ?

NIST National Institute of Standards and Technology